

# EN. 601.647/667

# INTRODUCTION TO HLT

# MACHINE TRANSLATION

---

KENTON MURRAY

9/28/2021

Slides in part thanks to:

Kevin Duh

David Chiang

Kevin Knight

# farok crrrok hihok yorok klok kantok ok-yurp .

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok **crrrok** hihok yorok klok kantok ok-yurp .

jjat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok <b>crrrok</b> hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok **hihok** yorok klok kantok ok-yurp .

jjat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok <b>hihok</b> ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok <b>hihok</b> yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok <b>hihok</b> mok . wat nnat forat arrat vat gat .



farok crrrok **hihok** yorok klok kantok ok-yurp .

jjat

arrat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok <b>hihok</b> ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok <b>hihok</b> yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok <b>hihok</b> mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok **yorok** klok kantok ok-yurp .

jjat

arrat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok <b>yorok</b> ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok <b>yorok</b> zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

farok crrrok hihok **yorok** klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok <b>yorok</b> ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok <b>yorok</b> zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok **cl**ok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok <b>cl</b> ok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	<p>lalok farok ororok lalok sprok izok enemok .          \          wat jjat bichat wat dat vat eneat .</p>
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	<p>lalok brok anak plok nok . iat lat pippat rrat nnat .</p>
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	<p>wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .</p>
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	<p>lalok mok nok yorok ghirok klok .          \          wat nnat gat mat bat hilat .</p>
wiwok farok izok stok . totat jjat quat cat .	<p>lalok nok crrrok hihok yorok zanzanok .          / /          wat nnat arrat mat zanzanat .</p>
lalok sprok izok jok stok . wat dat krat quat cat .	<p>lalok rarok nok izok hihok mok .          /          wat nnat forat arrat vat gat .</p>



# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok . at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok . iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok . totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok . at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok . totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok . wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok <b>nok</b> .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok <b>nok</b> izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok <b>nok</b> yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok <b>nok</b> crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok <b>nok</b> izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok <b>nok</b> .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok <b>nok</b> izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok <b>nok</b> yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok <b>nok</b> crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok <b>nok</b> izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>



# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok <b>ghirok</b> .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok <b>ghirok</b> klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok <b>ghirok</b> .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok <b>ghirok</b> klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok **cl**ok kantok ok-yurp .

jjat

arrat

mat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneat .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok <b>cl</b>ok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok **cl**ok kantok ok-yurp .

jjat

arrat

mat

bat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok <b>cl</b> ok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

bat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

bat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneak .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok crrrok hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

bat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>



# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

bat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneat .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok **crrrok** hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

bat

ok-voon ororok sprok . at-voon bichat dat .	lalok farok ororok lalok sprok izok enemok . wat jjat bichat wat dat vat eneat .
ok-drubel ok-voon anak plok sprok . at-drubel at-voon pippat rrat dat .	lalok brok anak plok nok . iat lat pippat rrat nnat .
erok sprok izok hihok ghirok . totat dat arrat vat hilat .	wiwok nok izok kantok ok-yurp . totat nnat quat oloat at-yurp .
ok-voon anak drok brok jok . at-voon krat pippat sat lat .	lalok mok nok yorok ghirok klok . wat nnat gat mat bat hilat .
wiwok farok izok stok . totat jjat quat cat .	lalok nok <b>crrrok</b> hihok yorok zanzanok . wat nnat arrat mat zanzanat .
lalok sprok izok jok stok . wat dat krat quat cat .	lalok rarok nok izok hihok mok . wat nnat forat arrat vat gat .

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

bat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok **kantok ok-yurp** .

jjat

arrat

mat

bat

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneat .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok <b>kantok ok-yurp</b> .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok **kantok ok-yurp** .

jjat

arrat

mat

bat

oloat

at-yurp

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneat .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok <b>kantok ok-yurp</b> .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# farok crrrok hihok yorok klok kantok ok-yurp .

jjat

arrat

mat

bat

oloat

at-yurp

<p>ok-voon ororok sprok .</p> <p>at-voon bichat dat .</p>	<p>lalok farok ororok lalok sprok izok enemok .</p> <p>wat jjat bichat wat dat vat eneak .</p>
<p>ok-drubel ok-voon anak plok sprok .</p> <p>at-drubel at-voon pippat rrat dat .</p>	<p>lalok brok anak plok nok .</p> <p>iat lat pippat rrat nnat .</p>
<p>erok sprok izok hihok ghirok .</p> <p>totat dat arrat vat hilat .</p>	<p>wiwok nok izok kantok ok-yurp .</p> <p>totat nnat quat oloat at-yurp .</p>
<p>ok-voon anak drok brok jok .</p> <p>at-voon krat pippat sat lat .</p>	<p>lalok mok nok yorok ghirok klok .</p> <p>wat nnat gat mat bat hilat .</p>
<p>wiwok farok izok stok .</p> <p>totat jjat quat cat .</p>	<p>lalok nok crrrok hihok yorok zanzanok .</p> <p>wat nnat arrat mat zanzanat .</p>
<p>lalok sprok izok jok stok .</p> <p>wat dat krat quat cat .</p>	<p>lalok rarok nok izok hihok mok .</p> <p>wat nnat forat arrat vat gat .</p>

# clients do not sell pharmaceuticals in Europe .

clientes                      no venden                      medicinas                      en                      Europa

Garcia and associates . Garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
Carlos Garcia has three associates . Carlos Garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in Europe . sus grupos estan en Europa .
Garcia has a company also . Garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequenos no son modernos .



*Warren Weaver*

**Warren Weaver,  
American scientist (1894-1978)**

When I look at an article in  
Russian, I say:  
"This is really written in English,  
but it has been coded in some  
strange symbols.  
I will now proceed to decode".



# Progress in MT



Warren Weaver's memo

1947

Founding of SYSTRAN.  
Development of Rule-based MT (RBMT)

1968

Seminal SMT paper from IBM

1993

DARPA TIDES, GALE, BOLT programs  
Open-source of Moses toolkit  
Development of Statistical MT (SMT)

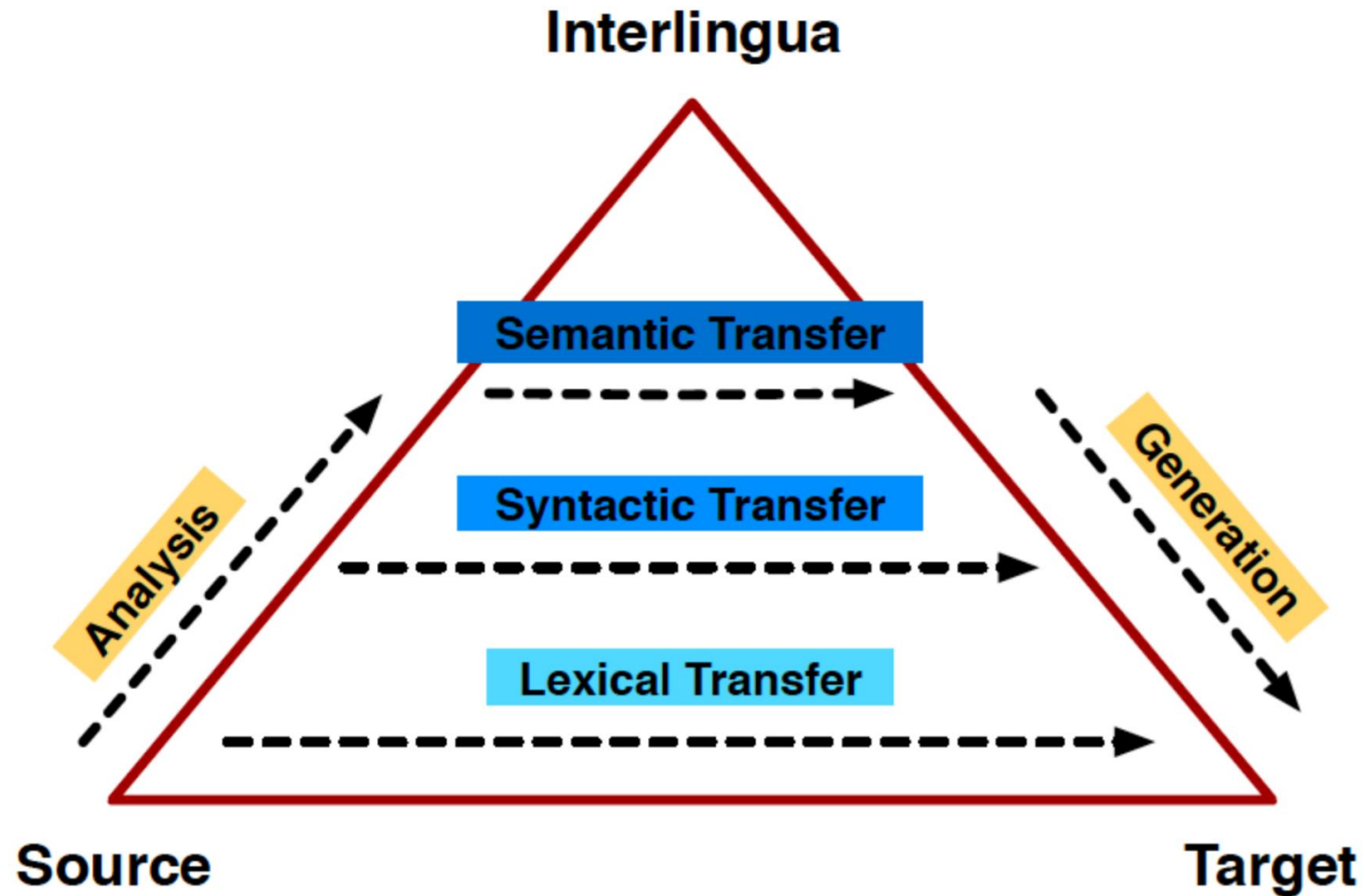
Early 2000s

2011-2012: Early deep learning success in speech/vision  
2015: Seminal NMT paper (RNN+attention)  
2016: Google announces NMT in production  
2017: New NMT architecture: Transformer

2010s-Present



# Vauquois Triangle



# RULE-BASED MACHINE TRANSLATION

---

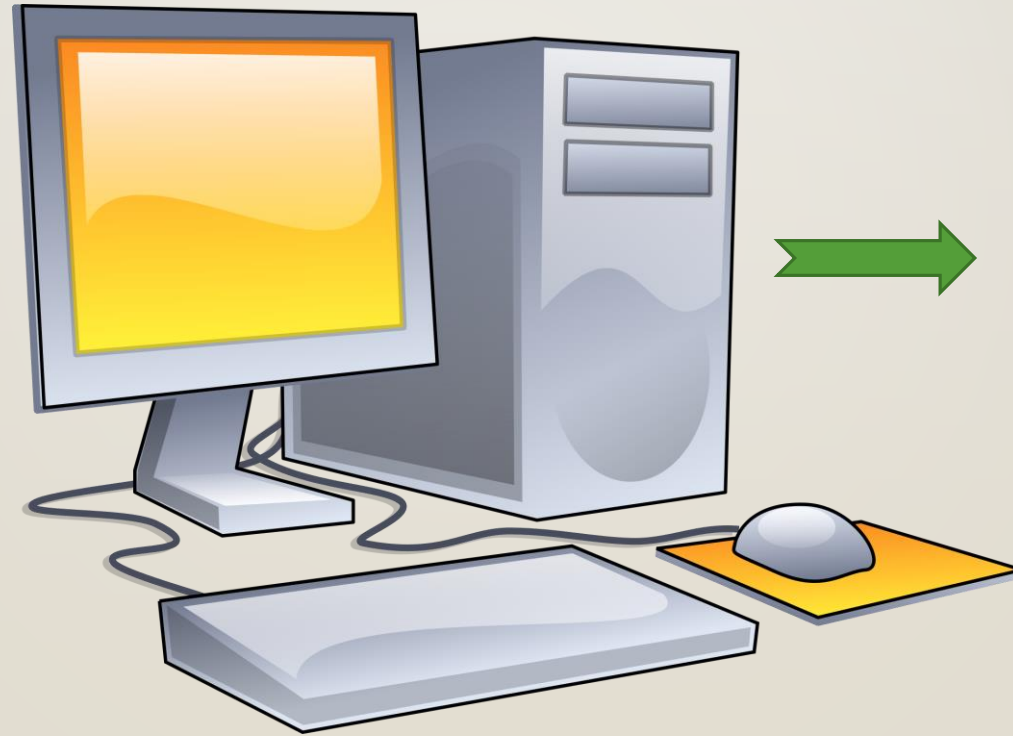
- Build Dictionaries
- Write Transformation Rules

```
"have" :=  
  
if  
  subject(animate)  
  and object(owned-by-subject)  
then  
  translate to "kade... aahe"  
if  
  subject(animate)  
  and object(kinship-with-subject)  
then  
  translate to "laa... aahe"  
if  
  subject(inanimate)  
then  
  translate to "madhye... aahe"
```

# STATISTICAL MACHINE TRANSLATION

---

كيف حالك



How are you?

# STATISTICAL MACHINE TRANSLATION

---



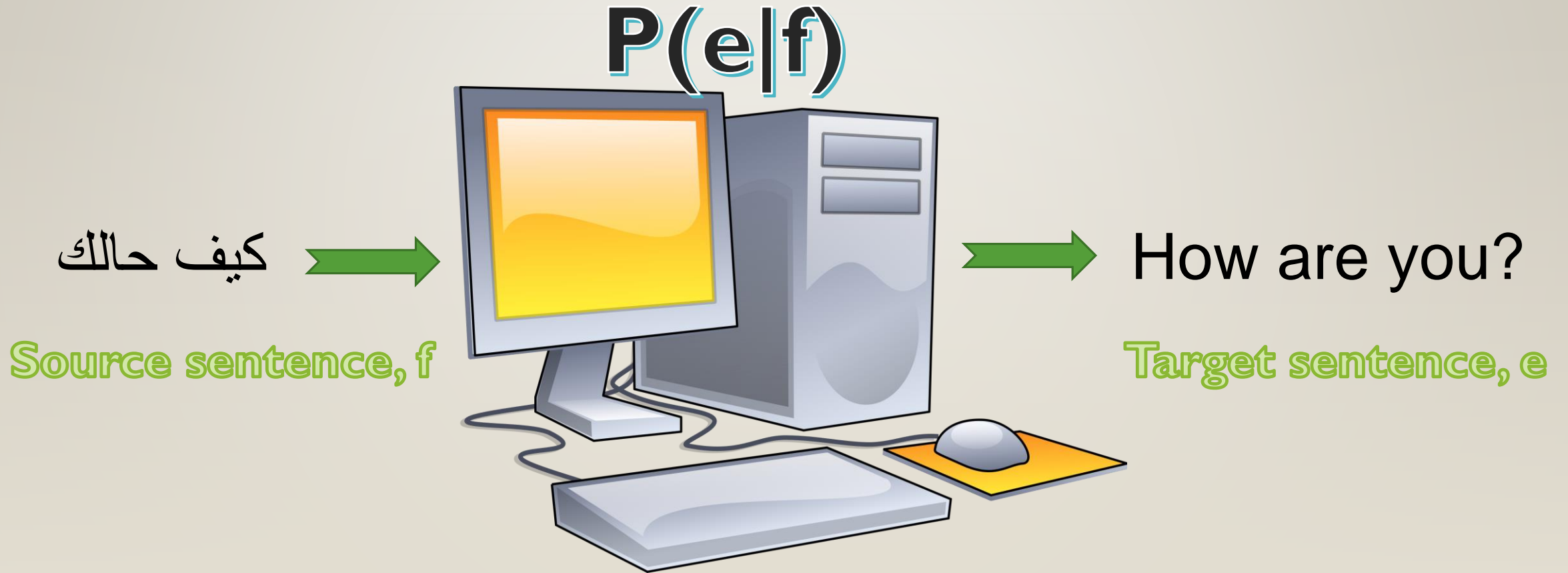
# STATISTICAL MACHINE TRANSLATION

---



# STATISTICAL MACHINE TRANSLATION

---



# STATISTICAL MACHINE TRANSLATION (SMT)

---

Data, Data, Data!



# STATISTICAL MACHINE TRANSLATION (SMT)

---

- Learn Dictionaries from Data

# STATISTICAL MACHINE TRANSLATION (SMT)

---

- Learn Dictionaries from Data “farok” → “jjat”

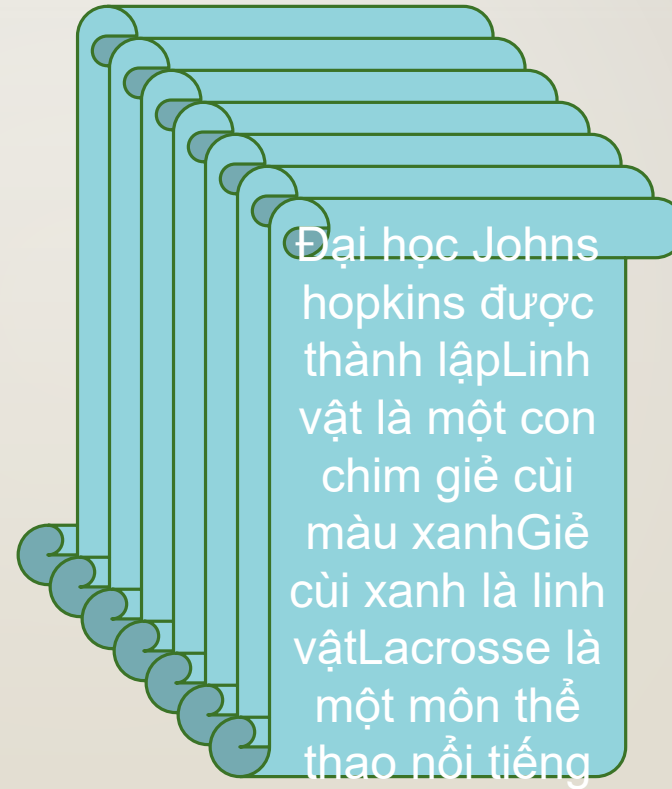
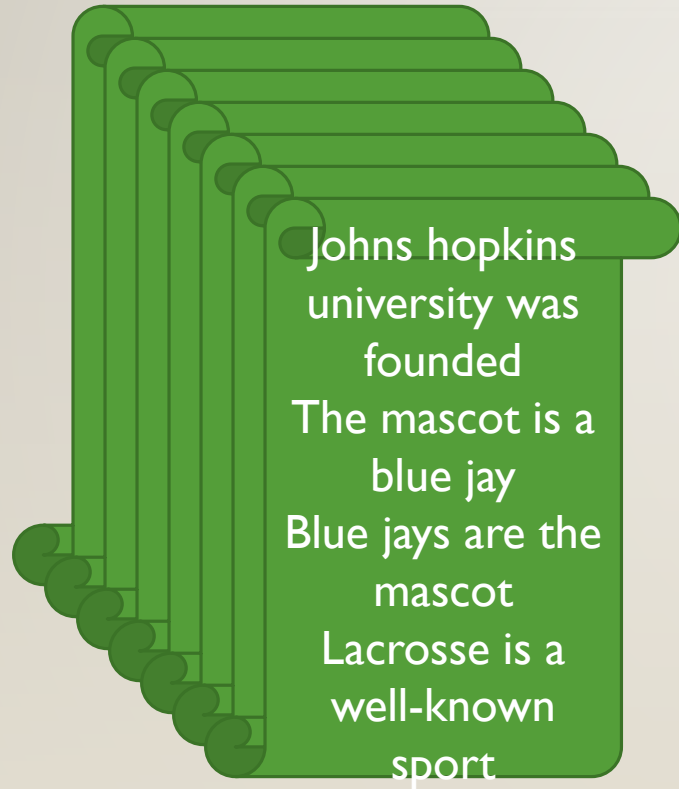
# STATISTICAL MACHINE TRANSLATION (SMT)

---

- Learn Dictionaries from Data “farok” → “jjat”
- Learn “Rules” from Data
- 1980 - 2015

# BITEXTS

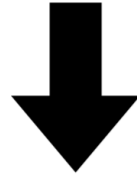
---



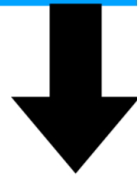
# MACHINE TRANSLATION (ABSTRACTION)

---

**There are 6000 languages in the world**



**Machine Translation (MT)  
System**



**世界には6000の言語があります**

# MACHINE TRANSLATION (SMT) ... SIMPLIFIED

There are 6000 languages in the world

↓ ↓ ↓ ↓ ↓  
あります 6000 言語 には 世界

TRANSLATION MODEL

↖ ↘ ↘ ↘ ↘  
世界 には 6000 の 言語 が あります

LANGUAGE MODEL & REORDERING MODEL

# SMT VERSUS NMT

---

Statistical MT	Neural MT
Input: Source Sentence	

# SMT VERSUS NMT

---

<b>Statistical MT</b>	<b>Neural MT</b>
Input: Source Sentence	Input: Source Sentence



# SMT VERSUS NMT


---

Statistical MT	Neural MT
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	

# SMT VERSUS NMT

---

<b>Statistical MT</b>	<b>Neural MT</b>
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	Output: Target Sentence



# SMT VERSUS NMT


---

<b>Statistical MT</b>	<b>Neural MT</b>
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	Output: Target Sentence
Automatically Learn from Bitext	

# SMT VERSUS NMT

---

<b>Statistical MT</b>	<b>Neural MT</b>
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	Output: Target Sentence
Automatically Learn from Bitext	Automatically Learn from Bitext




# SMT VERSUS NMT

Statistical MT	Neural MT
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	Output: Target Sentence
Automatically Learn from Bitext	Automatically Learn from Bitext
Probabilistic Translation Model	

# SMT VERSUS NMT

---

Statistical MT	Neural MT
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	Output: Target Sentence
Automatically Learn from Bitext	Automatically Learn from Bitext
Probabilistic Translation Model	
Probabilistic Reordering Model	



# SMT VERSUS NMT

Statistical MT	Neural MT
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	Output: Target Sentence
Automatically Learn from Bitext	Automatically Learn from Bitext
Probabilistic Translation Model	
Probabilistic Reordering Model	
Probabilistic Language Model	

# SMT VERSUS NMT

Statistical MT	Neural MT
Input: Source Sentence	Input: Source Sentence
Output: Target Sentence	Output: Target Sentence
Automatically Learn from Bitext	Automatically Learn from Bitext
Probabilistic Translation Model	One Neural Model (Probabilistic)
Probabilistic Reordering Model	
Probabilistic Language Model	



# NEURAL MACHINE TRANSLATION (NMT)

---

- Also a type of Statistical MT
- Represent words in high-dimensional, continuous, space
- $P(e|f)$

# NEURAL MACHINE TRANSLATION (NMT)

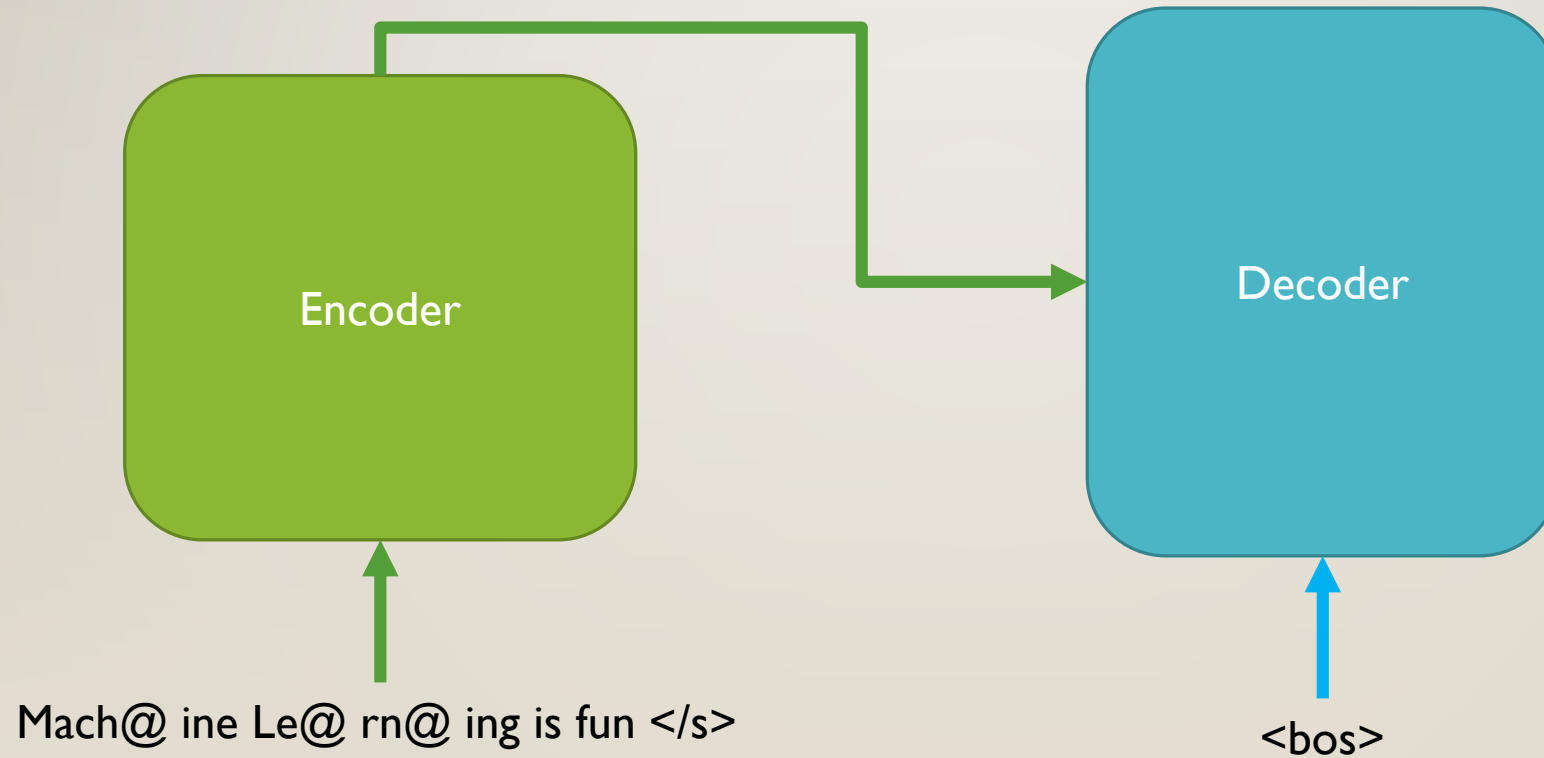
---



Mach@ ine Le@ rn@ ing is fun </s>

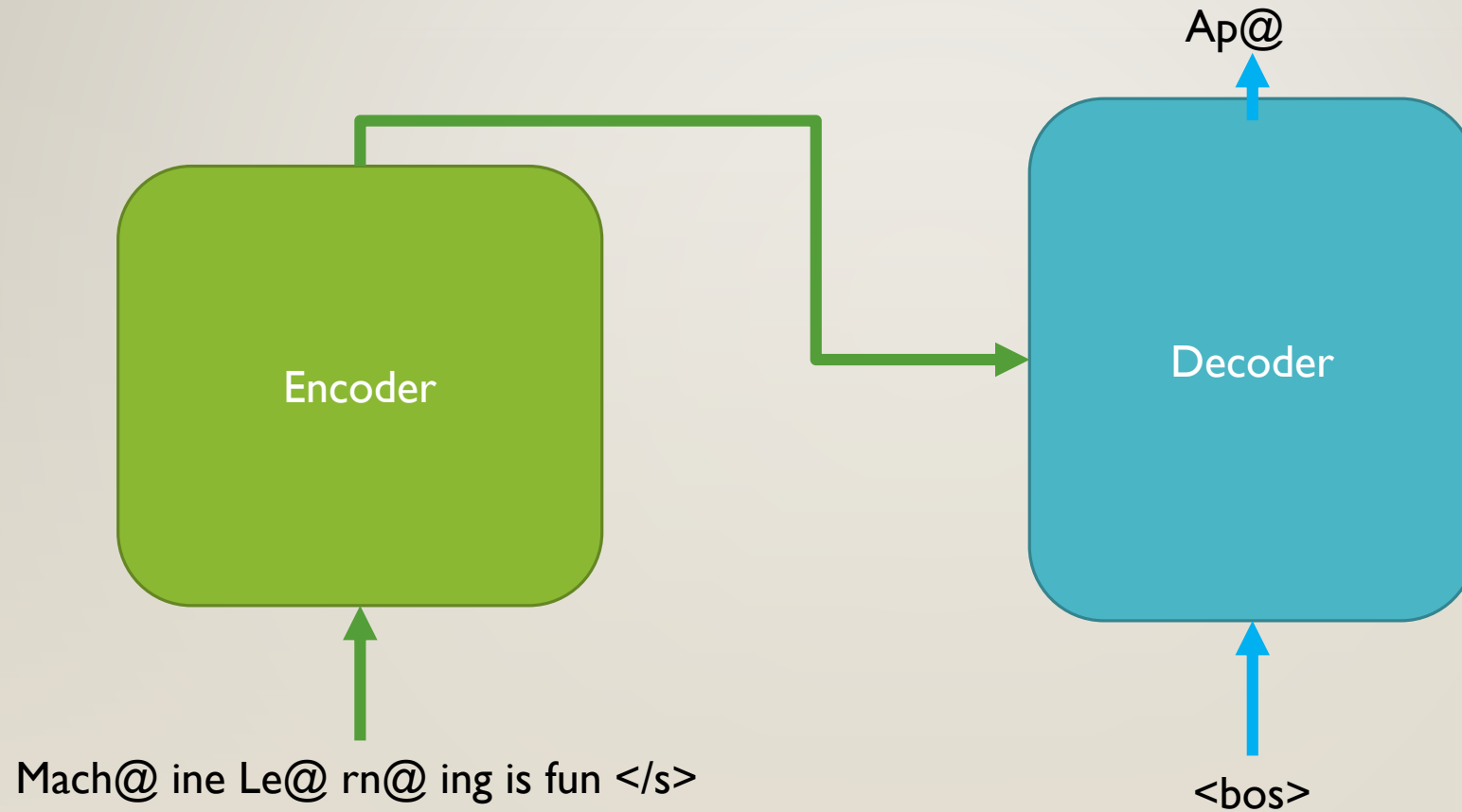
# NEURAL MACHINE TRANSLATION (NMT)

---



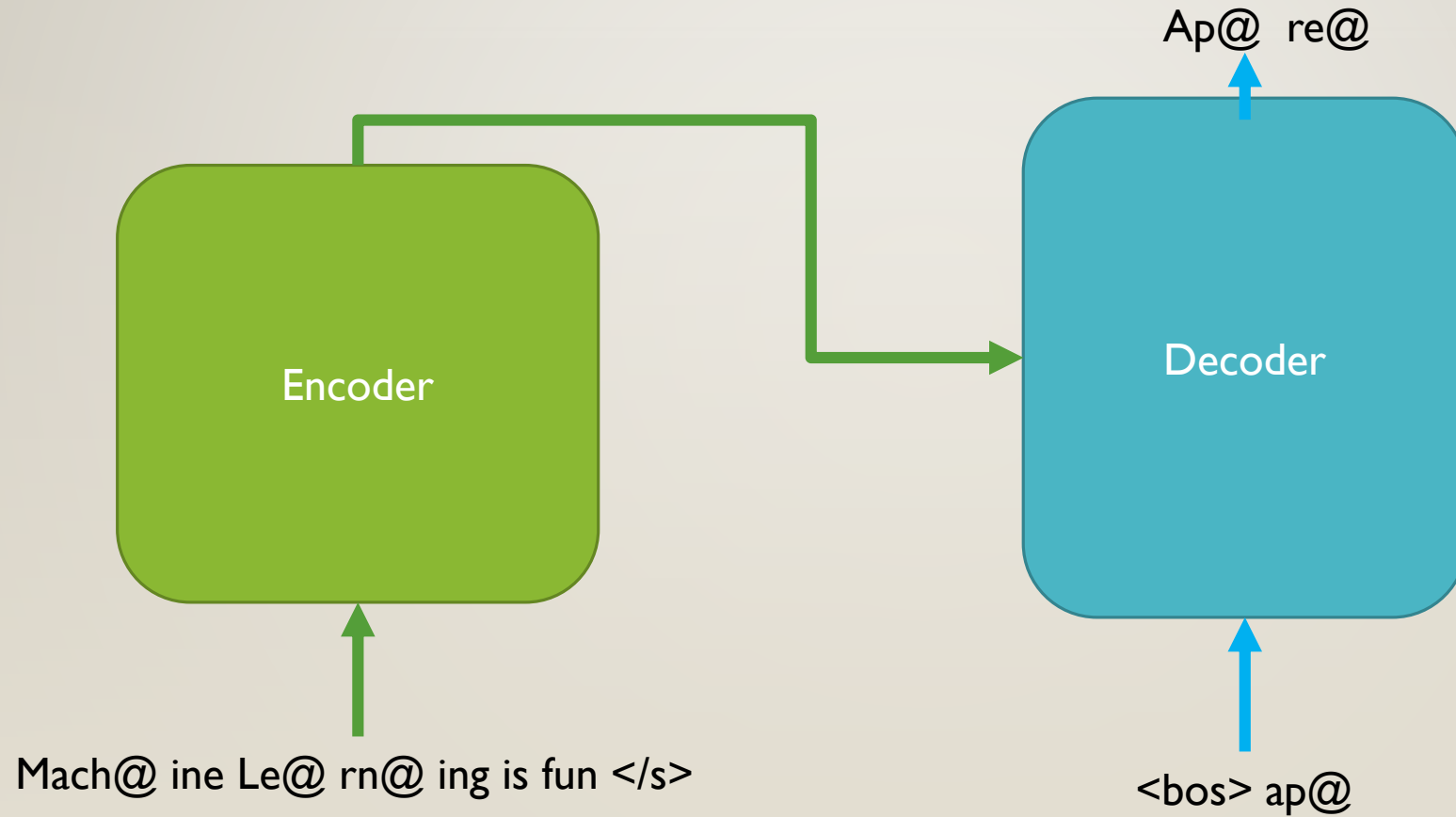
# NEURAL MACHINE TRANSLATION (NMT)

---



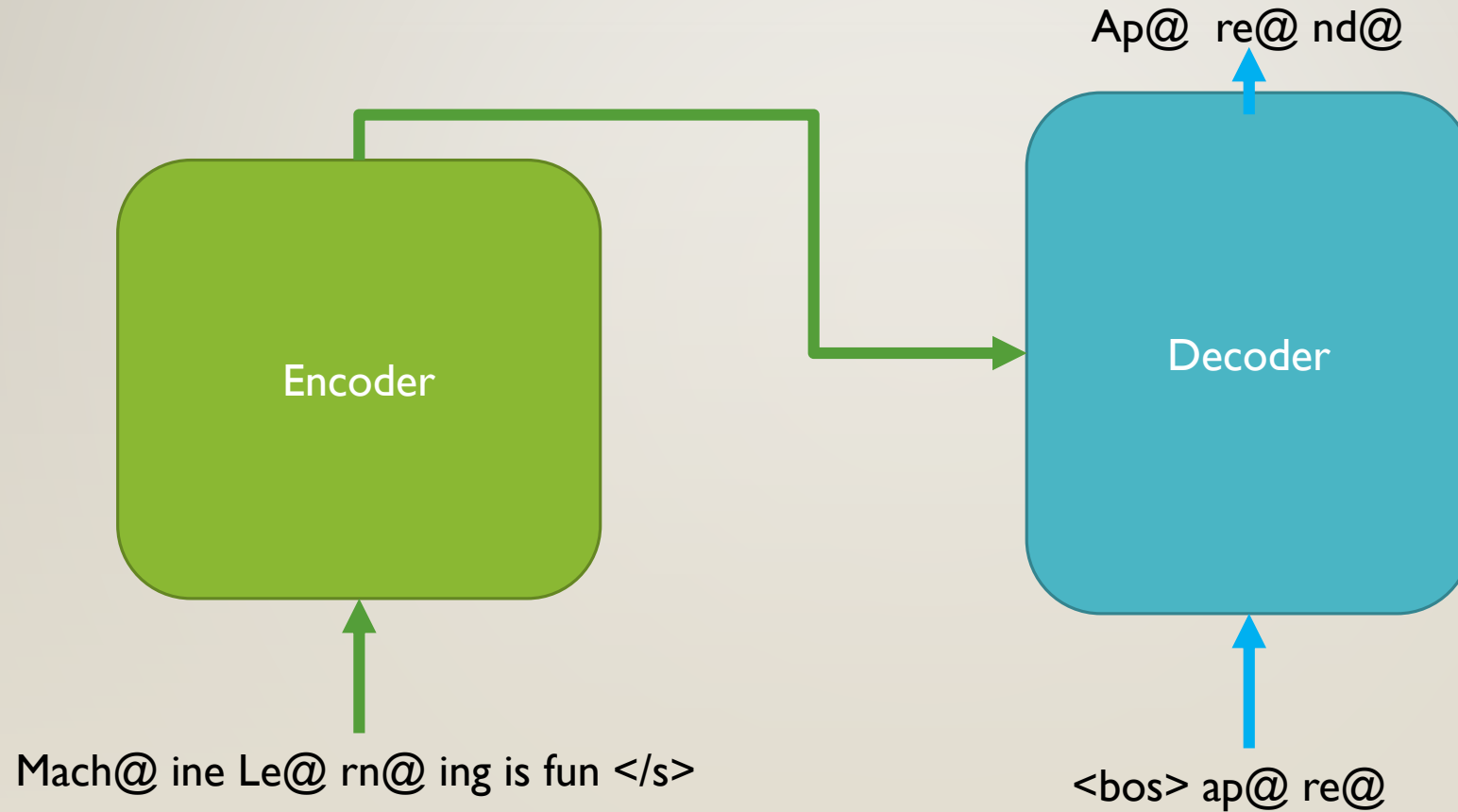
# NEURAL MACHINE TRANSLATION (NMT)

---



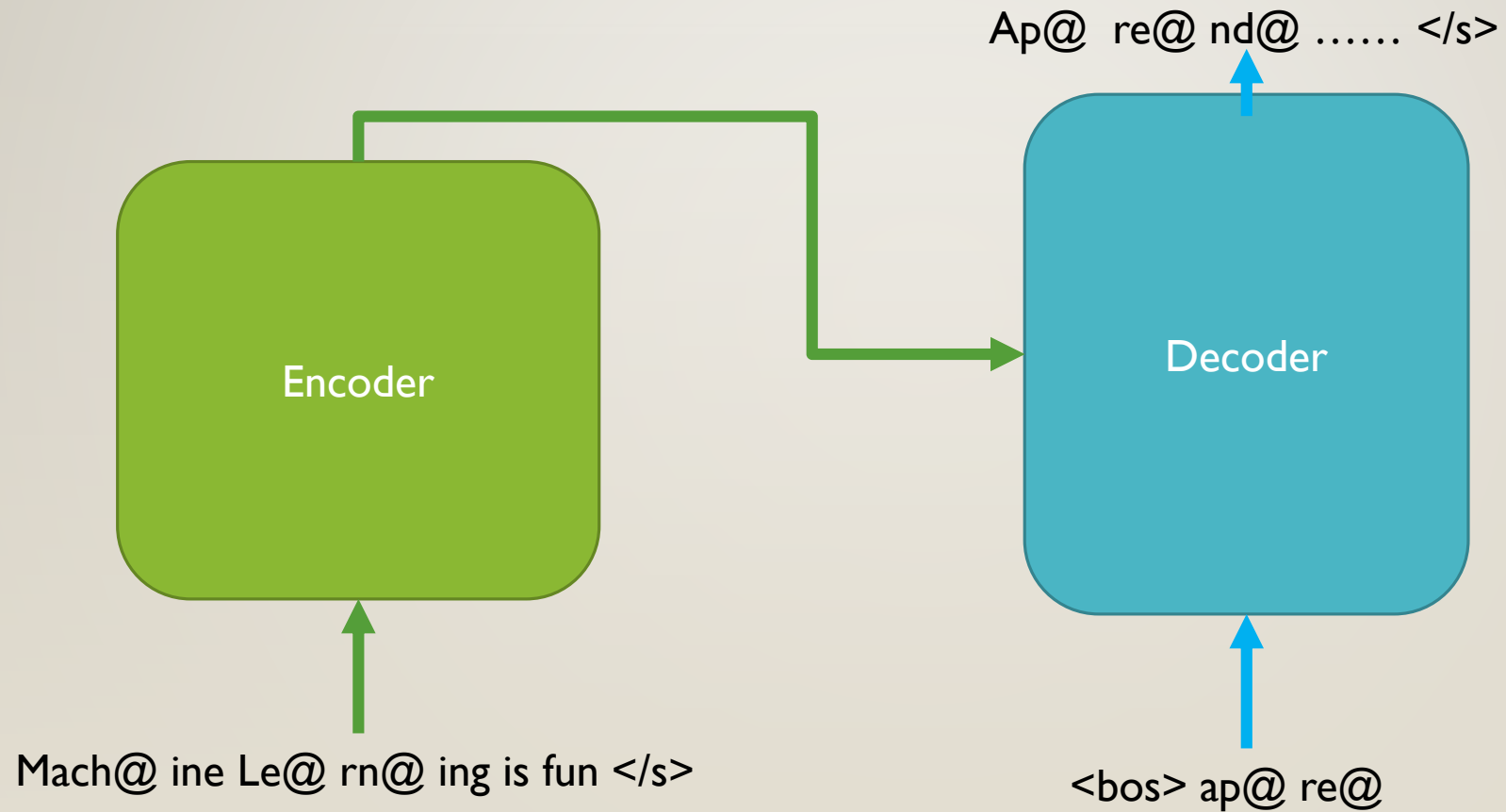
# NEURAL MACHINE TRANSLATION (NMT)

---

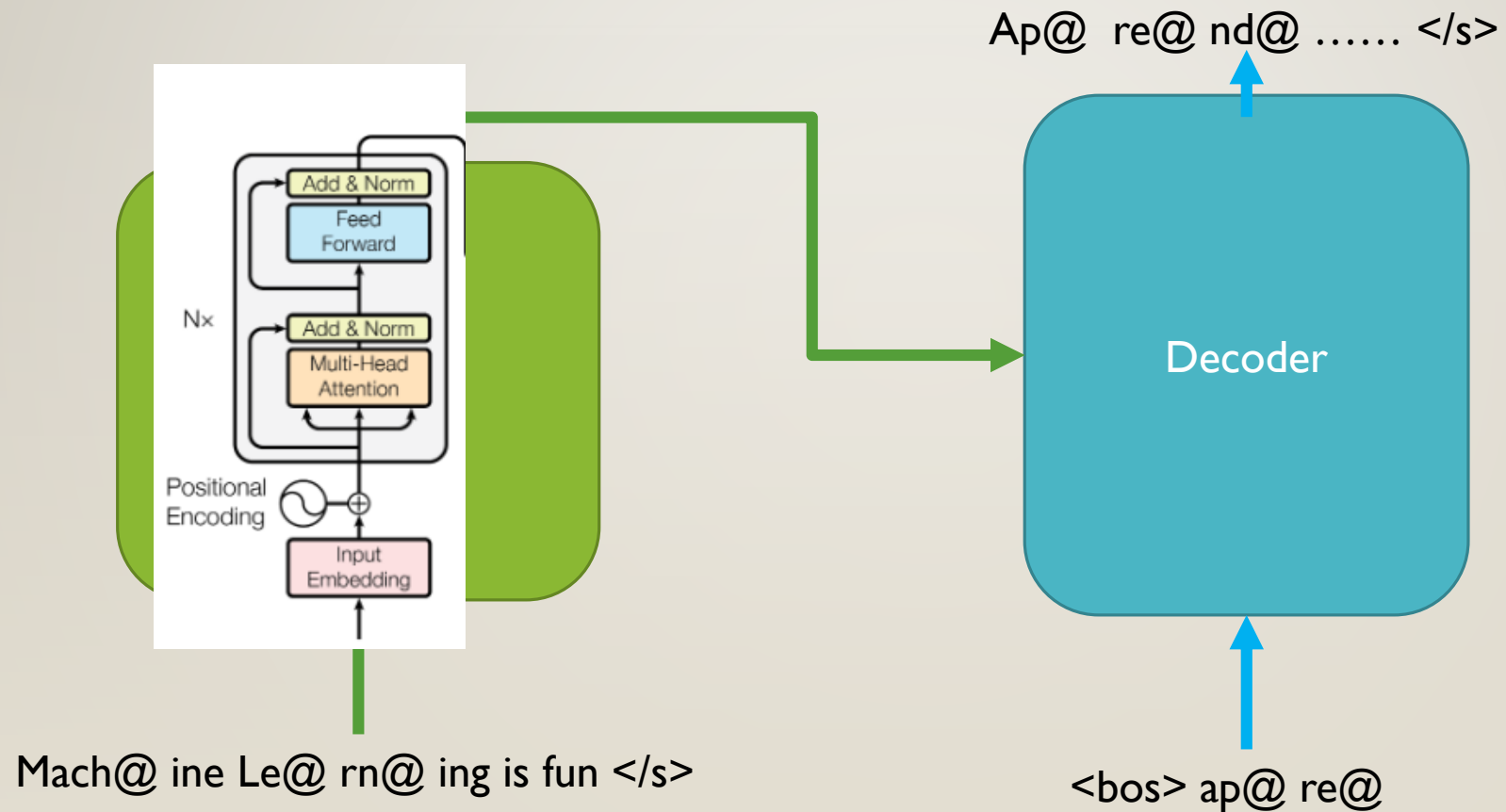


# NEURAL MACHINE TRANSLATION (NMT)

---

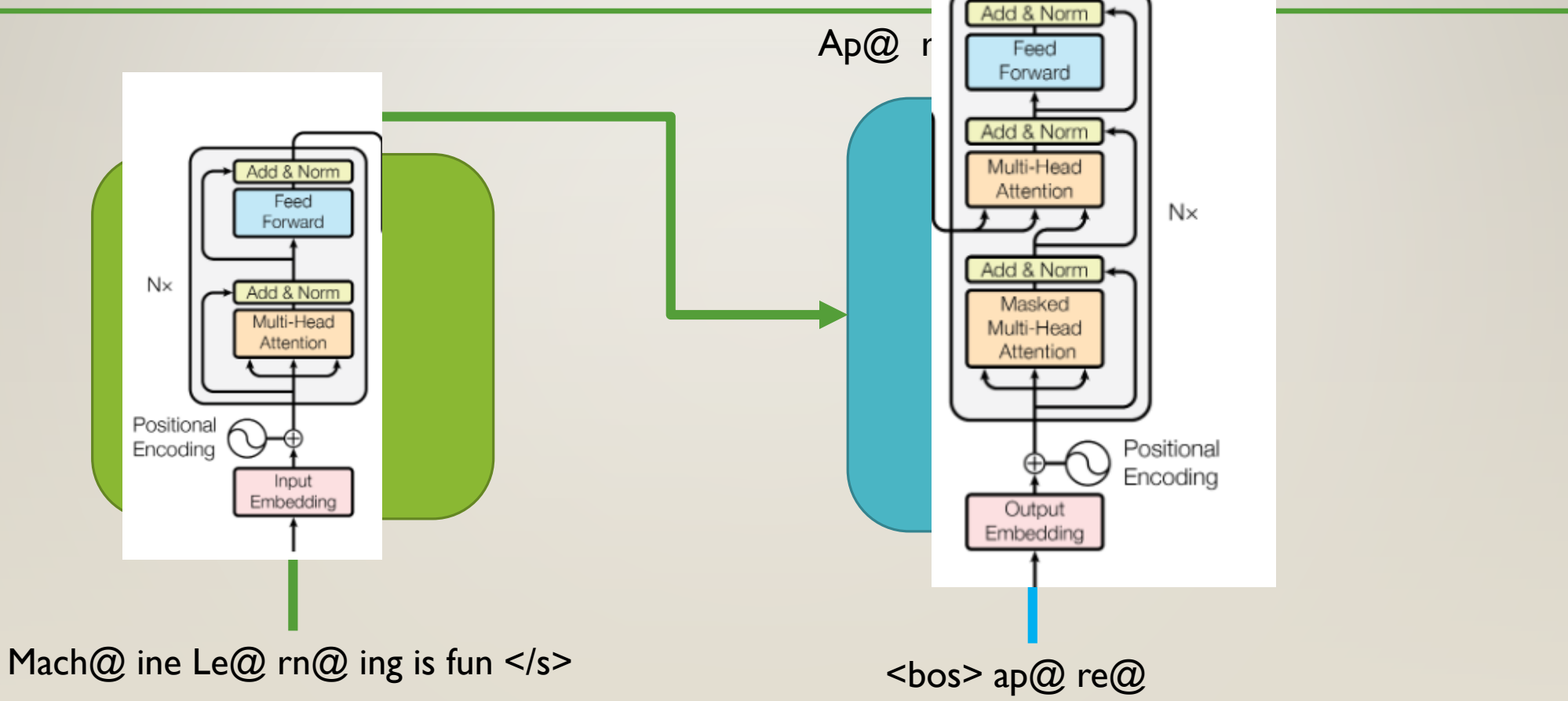


# NEURAL MACHINE TRANSLATION (NMT)





# NEURAL MACHINE TRANSLATION



Mach@ ine Le@ rn@ ing is fun </s>

<bos> ap@ re@

# VOCABULARY

---



# ONE-HOT VECTOR

---

- Words correspond to index in vector
- Fixed size

# ONE-HOT VECTOR

Dictionary

The And I Dog Johns ..... Me Cat

Johns Hopkins was founded in .....

0 0 0 0 | ..... 0 0



# CHARACTER LEVEL

---

- No OOVs
- Very long sequences

# BYTE PAIR ENCODING

---

- Subword Unit
- Based on a compression algorithm
- Start small, repeatedly combine

**peter piper picked a peck of pickled peppers**

---





p@ e@ t@ e@ r p@ i@ p@ e@ r p@ i@ c@ k@ e@ d a p@ e@ c@ k o@ f p@ i@ c@ k@ l@ e@ d p@ e@ p@ p@ e@ r@ s

---

<b>Vocabulary</b>	<b>Rules</b>
p t i k a f s	e r c d o l

p@ e@ t@ e@ r p@ i@ p@ e@ r p@ i@ c@ k@ e@ d a p@ e@ c@ k o@ f p@ i@ c@ k@ l@ e@ d p@ e@ p@ p@ e@ r@ s

---

Vocabulary		Rules
p	e	p@ e@ → pe@
t	r	
i	c	
k	d	
a	o	
f	l	
s	pe@	

pe@ t@ e@ r p@ i@ p@ e@ r p@ i@ c@ k@ e@ d a pe@ c@ k o@ f p@ i@ c@ k@ l@ e@ d pe@ p@ pe@ r@ s

---

Vocabulary		Rules
p	e	p@ e@ → pe@
t	r	p@ i@ → pi@
i	c	
k	d	
a	o	
f	l	
s	pe@	
pi@		

pe@ t@ e@ r pi@ p@ e@ r pi@ c@ k@ e@ d a pe@ c@ k o@ f pi@ c@ k@ l@ e@ d pe@ p@ pe@ r@ s

---

Vocabulary		Rules
p	e	p@ e@ → pe@
t	r	p@ i@ → pi@
i	c	pi@ c@ → pic@
k	d	
a	o	
f	l	
s	pe@	
pi@	pic@	

pe@ t@ e@ r pi@ p@ e@ r pic@ k@ e@d a pe@ c@ k o@ f pic@ k@ l@ e@d pe@ p@ pe@ r@ s

---

Vocabulary		Rules
p	e	p@ e@ → pe@
t	r	p@ i@ → pi@
i	c	pi@ c@ → pic@
k	d	e@ d → ed
a	o	
f	l	
s	pe@	
pi@	pic@	

peter piper picked a peck of pickled peppers

---

Vocabulary	Rules
p t i k a f s pi@ .....	p@ e@ → pe@ p@ i@ → pi@ pi@ c@ → pic@ e@ d → ed . . . .
of	o@ f → of

# HOW GOOD ARE OUR TRANSLATIONS?

---

- “Iyunivesithi yasekwa ngonyaka olandelayo ukusweleka kwakhe kwaye yanikezelwa kunyana wabo okuphela kwakhe.”
- The university was founded in the year following his death and was dedicated to their only son.
- The university was following his death and was dedicated to their only son.
- Machine Translation always works perfectly.

# MT EVALUATION

---

- Human Evaluation (Expensive, but best)
- Automatic (Cheap, can be correlated)



# BLEU SCORES

---

- Modified  $n$ -gram precision
- BiLingual Evaluation Understudy

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus.

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around.



# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16			

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16			

# BLEU SCORES

---

**Gold:** The **Blue Jays** are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The **Blue Jays** are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16			

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15		

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of [Johns Hopkins University](#) and can be seen around campus

**HYP:** The Blue Jays are mascot of [Johns Hopkins University](#) University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	/14	

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	



# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	/13

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins **University and can be** seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University **University and can be** seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	/13

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	/13

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and **can be seen around** campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and **can be seen around**

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	/13

# BLEU SCORES

---

**Gold:** The Blue Jays are the mascot of Johns Hopkins University and can be seen around campus

**HYP:** The Blue Jays are mascot of Johns Hopkins University University and can be seen around

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	4/13

# BLEU SCORES

---

$$BLEU = \min \left( 1, e^{1 - \frac{\text{len}(\text{gold})}{\text{len}(\text{hyp})}} \right) \sqrt[4]{\prod_{i=1}^4 \text{precision}(i)}$$

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	4/13

# BLEU SCORES

---

$$BLEU = \min \left( 1, e^{1 - \frac{\text{len}(\text{gold})}{\text{len}(\text{hyp})}} \right) \sqrt[4]{\prod_{i=1}^4 \text{precision}(i)}$$

Brevity Penalty

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	3/13

# BLEU SCORES

---

$$BLEU = \min \left( 1, e^{1 - \frac{\text{len}(\text{gold})}{\text{len}(\text{hyp})}} \right) \sqrt[4]{\prod_{i=1}^4 \text{precision}(i)}$$
$$BLEU = \min \left( 1, e^{1 - \frac{16}{15}} \right) \sqrt[4]{0.138}$$

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	3/13



# BLEU SCORES

---

$$BLEU = \min \left( 1, e^{1 - \frac{\text{len}(\text{gold})}{\text{len}(\text{hyp})}} \right) \sqrt[4]{\prod_{i=1}^4 \text{precision}(i)}$$
$$BLEU = 0.94 \sqrt[4]{0.138}$$

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	3/13

# BLEU SCORES

---

$$BLEU = \min \left( 1, e^{1 - \frac{\text{len}(\text{gold})}{\text{len}(\text{hyp})}} \right) \sqrt[4]{\prod_{i=1}^4 \text{precision}(i)}$$

$$BLEU = 0.94 * 0.609$$

$$BLEU = 0.57$$

Unigrams	Bigrams	Trigrams	4-grams
14/16	12/15	9/14	3/13

# BLEU SCORES

---

- Calculate over entire test set (not one sentence)
- $< 10$  .... Pretty useless
- $10 - 20$  ... can get some meaning
- $20 - 30$  ... looks decent
- $> 30$  starts getting pretty good

# BLEU SCORES

---

- The large house
- A big mansion

# FAIRSEQ

---

- PyTorch
- <https://github.com/pytorch/fairseq>

